Use of locally weighted scatterplot smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds

Pintus, E. , Sorbolini, S. , Albera, A. , Gaspa, G. , Dimauro, C. , Steri, R. , Marras, G. and Macciotta, N. P.

Use of locally weighted scatterplot smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds.

Anim Genet, 45: 1-11, 2014,
doi:10.1111/age.12076

Use of Locally Weighted Scatterplot Smoothing (LOWESS) regression to study selection signatures in Piedmontese and Italian Brown cattle breeds

Elia Pintus[1], Silvia Sorbolini[1], Andrea Albera[2], Giustino Gaspa[1], Corrado Dimauro[1], Roberto Steri[1], Gabriele Marras[1], Nicolò Pietro Paolo Macciotta[*1]

Address: [1]Dipartimento di Agraria, Sezione di Scienze Zootecniche Università degli Studi di Sassari, 07100 Sassari, Italy, [2]Associazione Nazionale Allevatori Bovini di Razza Piemontese (ANABORAPI), 12061 Carrù (CN), Italy.

Email: Elia Pintus – epintus@uniss.it; Silvia Sorbolini - ssorbolini@uniss.it; Andrea Albera - info@anaborapi.it; Giustino Gaspa – gigaspa@uniss.it; Corrado Dimauro - dimauro@uniss.it; Roberto Steri - rsteri@uniss.it; Gabriele Marras - gbmarras@uniss.it; Nicolò Pietro Paolo Macciotta[*] - macciott@uniss.it;

[*]Corresponding author

Summary

Selection is the major force affecting local levels of genetic variation in species. The availability of dense marker maps offers new opportunities for a detailed understanding of genetic diversity distribution across the animal genome. Over the last fifty years, cattle breeds have been subjected to intense artificial selection. Consequently, regions controlling traits of economic importance are expected to exhibit selection signatures. The fixation index (F_{st}) is an estimate of population differentiation, based on genetic polymorphism data and it is calculated using the relationship between inbreeding and heterozygosity. In the present study, the locally weighted

scatterplot smoothing regression (LOWESS) and a Control Chart approach were used to investigate selection signatures in two cattle breeds with different production aptitude (dairy and beef).

F_{st} was calculated for 42,514 SNPs marker loci distributed across the genome in 749 Italian Brown and 364 Piedmontese bulls respectively. The statistical significance of F_{st} values was assessed using a Control Chart. The LOWESS technique was efficient in removing noise from the raw data and was able to highlight selection signatures in chromosomes known to harbour genes affecting dairy and beef traits. Examples are, the peaks detected for BTA2 in the region where the *myostatin* gene is located and for BTA6 in the region harbouring the *ABCG2* locus. Moreover, several loci not previously reported in cattle studies were detected.

Key words: SNPs, F_{st}, LOWESS, cattle breeds

Introduction

The study of the genetic basis of differences among animal populations is a hot topic of animal genetics. The quantification of allelic richness and the evaluation of their association with phenotypes represent tools for the safeguard and the management of local populations. Moreover, identification of genomic regions involved in phenotypic differences between individuals provide useful knowledge for gene assisted selection programmes.

Specialized breeds have been subjected to intense selection. A main consequence has been the progressive erosion of local levels of genetic variation that may have compromised the ability to challenge environmental factors (Mäki-Tanila et al. 2010). Thus a genetic comparison between selected and autochthonous populations may allow for the identification of genomic regions involved in the control of fitness traits. On the other hand, studies involving highly selected breeds with different production aptitudes, as the case of dairy and beef cattle, can provide an exciting opportunity for studying signatures of selective breeding (Hayes et al. 2008a; Qanbari et al. 2010).

Actually, little is known about the effects of intensive, directional and prolonged selection on genome sub-structure of domestics species.

In population genetics, the identification of a locus target of selection is based on the existence of a reduction in nucleotidic diversity, or on an increase linkage disequilibrium (LD) and/or a changed allele frequency (Doebley et al. 2006). Currently, different statistical methods are used for the detection of selection signatures. Many of them are based on the comparison of allele frequencies or haplotype structure (for a review see Biswas and Akey 2006). The most commonly used metrics are the r^2 for measuring LD and the fixation index F_{st} (Weir and Cockerham, 1984). A quite recent approach, extensively studied in human populations, is based on the detection of runs of homozygosity (ROH), defined as uninterrupted stretch of homozygous genotypes (Gibson et al., 2006). The extent and frequency of ROHs can be used as an indication of past or recent inbreeding (Khatkar et al., 2010; Purfield et al., 2012; Ferencakovic et al., 2012). However, for many of these methods it is difficult to develop a proper statistical test. This is particularly true when searching for selective signatures within a single population.

High throughput platforms able to simultaneously genotype for many thousands of SNP offer a powerful tool for the assessment of the genetic diversity across the genome (Andersson and Georges 2004). Genome Wide Analysis (GWA) have been performed to clarify the role of selection and drift in the evolutionary processes (Biswas and Akey 2006). Several recent studies have proposed the hitch-hiking mapping approach for identification of target of positive selection. The basic assumption is that the substitution of favourable allele at one site results in a reduction of variability at closely linked sites and lead to the allele fixation in a population (Przeworski et al. 2005). Actually, the abundance of SNP throughout the genome makes them particularly suitable in the detection of such selective sweeps (Andersson and Georges 2004).

However, such a huge amount of information has become rather problematic to interpret. A

major issue is represented by the great variability of the signal pattern (for example heterozygosity or other related statistics as F_{st}) along the chromosome. An usual empirical practice to smooth data is to work on average values of sliding windows of predetermined size (Weir et al. 2005; Barendse et al. 2009). A common problem is represented by the development of a suitable statistical test able to assess whether an index of the genetic difference between two populations can be considered significant. Different approaches have been proposed to assess F_{st} statistical significance, as the calculation of q-values for Kernel-smoothed values (Flori et al., 2009), permutation test based on binomial distribution of the SNP allelic frequencies (Stella et al., 2010) or the setting of a threshold of one standard deviation from the mean for smoothed F_{st} values (Kijas et al., 2012).

In the present work, an approach for studying selection signatures in two Italian cattle breeds with different production aptitude, Italian Brown and Piedmontese (dairy and beef, respectively), is proposed. In particular, a local regression is used to smooth raw F_{st} data and a Control Chart is applied to predicted data for identifying significant values. The method is challenged to identify genes that have been reported to be involved in the genetic determinism of dairy and beef traits in cattle.

Materials and Methods

A sample of 749 Italian Brown and 364 Piedmontese bulls was considered in the study. Animals were genotyped with the Illumina Bovine bead-chip containing 54,001 SNP (http://www.illumina.com). Only SNP located on the 29 autosomes in the Btau4.0 build of the Bovine Genome assembly were considered. Quality control was performed on the genotypes. SNP were removed if: monomorphic in both breeds; had a percentage of missing data higher than 2.5% had a minor allele frequency lower than 1%. After data editing, 42,514 markers were retained for the study. Missing data were replaced with the most frequent allele at that specific locus for each

97 breed.

98 Allele frequencies, observed and expected heterozygosity were calculated for each breed.
99 Total allelic frequencies for each locus, f_p and f_q, considering all animals as a single population were
100 calculated as:

101 $f_p = [f_B \cdot (2 \cdot n_B) + f_P \cdot (2 \cdot n_P)] / (n_B + n_P);$

102 Where f_B and n_B are frequencies of alleles and number of individuals in Brown, and f_P and
103 n_P are frequencies of alleles and number of individuals in Piedmontese.

104 $f_q = 1 - f_p$

105 Then, expected heterozygosity in populations (Hs) and overall (Ht) were calculated. Finally,
106 F_{st} was calculated according to Weir and Cockerham (1984) as:

107 F_{st} =Ht-Hs/Ht

108 In order to smooth F_{st} pattern, data were fitted with a Locally Weighted Scatterplot
109 Smoothing (LOWESS) regression using the PROC LOWESS of SAS/STAT software version 9.2
110 (SAS Institute, Inc; Cary NC) (Cleveland 1979). The LOWESS has been used in genetics for
111 smoothing model R^2 in the statistical analysis of molecular marker data (Questa-Marcos et al.,
112 2010). In the LOWESS, the space of the independent variable is fragmented into different intervals
113 for which separate regressions are fitted. The aim of the method is to remove noise from raw data
114 and to clarify graphical presentations. A critical point in fitting LOWESS is the identification of a
115 suitable dimension of the data interval to be included in the analysis. In other words, if x is the
116 number of adjacent points to be used in the estimation procedure of a set of n data, each region
117 contains a fraction of points given by x/n. This ratio is defined as the smoothing parameter S of the
118 LOWESS regression. As S increases, the fitted line will be smoother until $S = 1$ that corresponds to
119 a single line (i.e. the standard linear regression). Consequently, the goodness of fit depends strongly
120 on the smoothing parameter used (Cohen 1999).

In general, the number of markers considered in the local regression was different across chromosomes, being directly related to their length. Therefore the use of the same *S* parameter in all chromosomes could not be feasible. In the present work, a smoothing parameter corresponding to an interval of 20 SNPs for each separated regression gave the best results. The different smoothing parameters used for each chromosome are reported in Table 1 (supplemental material).

In order to identify F_{st} values different from the average pattern that could be evidence of selection signatures, LOWESS smoothed F_{st} were analysed with a Control Chart approach. This methodology aims at checking a process and its variability and it can be used to identify sources of variation. In the specific case of the present study, the goal was to partition F_{st} variation into a component due to selection, that causes a drop of heterozygosity, and a remaining random variation along the chromosome. Control Charts are graphically displayed as stream of data falling within control limits. Data exceeding these limits are flagged as outlier signals. A similar approach has been used by Kijas et al. (2012) for identifying selection signatures in sheep breeds. A Control Chart approach has been recently used to identify sites of preferential location of genetic variation in *Mycobacterium tuberculosis* (Das et al., 2012). In the present study, smoothed F_{st} values were plotted against their position along the chromosome. Limits of the Control Chart were set at 3 standard deviations from the mean.

In order to compare the results with an assessed methodology for studying selection signatures, F_{st} values were also smoothed with a sliding windows approach. The genome was divided into windows and average F_{st} values for each interval were calculated. Sliding windows are a graphical method widely used for detect genomic regions under positive or balancing selection (Hayes et al., 2008a; Stella et al., 2010). In the present study, the size of the window was fixed at 20 SNP each (i.e. of the same size of those used in the LOWESS smoothing). The two methods were compared by examining patterns of smoothed F_{st} signals.

Annotated genes in genomic regions corresponding to peaks exceeding Control Chart limits were derived from the UCSC Genome Broswer Gateway (http://genome.ucsc.edu/). Intervals of 500Kbp (0,25 Mbp upstream and downstream the significant region) were considered.

Results and Discussion

The comparison of chromosome average heterozygosity (Hobs) between the two breeds highlights lower values for the Italian Brown (average difference of 0.04) (Figure 1). The largest difference was found for BTA6 (0.07) the smallest for BTA2 (0.02). Differences in heterozygosity between cattle breeds have been reported by other authors (Ciampolini et al. 1995; Cañón et al. 2001).

In general, non smoothed SNP F_{st} values were characterized by some well defined peaks that could be evidence of divergent selection, and by a large background of low to moderate values indicating random noise. The largest number of high raw F_{st} values (n= 17, between 0.4 and 0.9) was detected on BTA6, the smallest (n= 1 with F_{st} value= 0.26) on BTA23 (Figures 2a and b, respectively). Few raw F_{st} signals (both in frequency and magnitude) were also detected on BTA28 and 29 (supplemental material). The pattern of raw F_{st} data for BTA6 was more regular compared to BTA23. This result may be interpreted as a consequence of the hitchhiking effect, because a reduction of heterozygosity (selective sweep) affect polymorphism of both individual and associated loci (Stephan et al., 2006).

The LOWESS correction resulted in a better definition of highest peaks, even if with an expected reduction in scale due to the regression (supplemental material). Moreover, other peaks of moderate height have been disentangled from the background noise of raw F_{st} data. As an example, LOWESS corrected F_{st} values for BTA6 and BTA23 are reported in Figures 2c and d, respectively. It can be clearly seen that the smoothing procedure enhanced clustered peaks whereas isolated

169 signals were regressed towards lower values. This behaviour was observed for the whole genome
170 (supplemental material).
171 Compared to other methods currently used for studying selection signatures, such a relative
172 simplicity could be interpreted as a sign of weakness. Actually, the LOWESS regression is a robust
173 non parametric method, does not relies on strong assumptions on data distribution and it could be
174 considered as a sort "of vertical sliding windows" (Jacoby, 2000). Such a property was evident also
175 in the comparison with the sliding windows approach performed in this study (Figure 3): the
176 LOWESS was actually able to yield more defined and clear signals. The enhancement of cluster of
177 peaks and the lowering of isolated signals are evidence of robustness of the method that is not
178 affected by the variation of a single marker. This feature is particularly useful for fitting the hitch-
179 hiking effect that occurs in the surroundings of a selectively favourable mutation (Maynard Smith
180 and Haigh, 1974).
181 The control chart analysis detected a total of 98 outliers on the whole genome. BTA6, 8 and
182 15 showed the largest number of signals (8, 8 and 7 respectively). On BTA11, 12, 17 and 22 a
183 single signal was detected, whereas no peaks were found on BTA23, 25 and 29. These figures are
184 lower than those reported by Stella et al. (2010) that, on a large number of breeds, found 699
185 different putative selection signatures on the whole genome. However, Flori et al. (2009) using
186 smoothed F_{st} across three different dairy or dual purpose breeds identified a total of 13 significative
187 regions under selection distributed on seven distinct chromosomes. Some of these regions
188 correspond to those detected in the present work. An example are peaks of smoothed F_{st} that have
189 been found on BTA6 regions where *LAP3* (*leucine aminopeptidase 3* at 37,871,423-37,896,860 bp)
190 and *LCORL* (*ligand dependent nuclear receptor corepressor-like* at 38,137,617-38,288,047 bp) loci
191 map.
192 Five peaks distributed throughout the chromosome were detected BTA19. A total of 66

193 different annotated loci for the corresponding genomic regions were retrieved from UCSC Genome
194 Browser Gateway data base. This is the highest number of genes per chromosome found in the
195 present analysis. This result is in agreement with the study of Band et al. (2000), that reported a
196 significantly larger number of mapped genes for BTA19 compared to the other autosomes.

197 As far as the use of the Control Chart for testing outliers is concerned, the way confidence
198 limits are set implies an assumption of normality for data distribution. Actually F_{st} often shows a
199 heavily skewed distribution (Deng et al., 2007). However, a way to deal with this problem is to
200 divide data into subgroups and then use their averages which could be considered approximately
201 normally distributed (Morrison, 2008). Such an approach is similar to what has been done in the
202 present work, where actually intervals of F_{st} data were considered for the LOWESS smoothing.
203 Recently, to investigate levels of genetic diversity and to characterise the role of domestication and
204 selection on the sheep genome, Kijas and co-workers (2012) performed a genome-wide analysis
205 using smoothed SNP-specific F_{st} plotted for values on excess of one standard deviation from the
206 mean. In any case, being straightforward to interpret without specific statistical background and
207 simple to update, the Control Chart approach has been widely used in genetics, medicine and other
208 fields of applied biology (Westgard et al., 1981; Coskun et al., 2008; Das et al., 2012).

209 The reliability of the proposed method was confirmed by smoothed F_{st} values that exceeded
210 Control Chart limits in regions of the genome where genes known to affect productive traits are
211 located. An evident example was the highest peak detected at about 37 Mbp on BTA6 (Figure 2e).
212 It was the largest smoothed F_{st} predicted value (0.30) observed across the whole genome in the
213 present study. Some genes known to affect milk production traits have been mapped in this region.
214 Examples are *Family with sequence similarity 13 member A* (*FAM13A1*) (36,740,247-36,843,133
215 bp) (Cohen et al. 2004), *ATP-binding Cassette, sub-family G (WHITE), member 2* (*ABCG2*)
216 (37,342,201-37,433,870 bp), *secreted phosphoprotein 1* (*SPP1*) (37,511,672-37,511,830 bp) and

217 *peroxisome proliferator-activated receptor gamma, coactivator 1 alpha* (*PPARGC1A)* (44,797,216-
218 44,935,623 bp) (Cohen-Zinder et al. 2005; Ron and Weller 2007; Sheehy et al. 2009). On the other
219 hand, no LOWESS predicted Fst peaks were detected on BTA23 (Figure 2f).
220 A further example is represented by a peak exceeding the chart limits that was detected
221 between 6,5-7,5 Mbp on BTA2. It is well known that *myostatin* (*MSTN*) locus that controls double
222 muscling phenotype in cattle is located in position 2q14-q15 between 6,532,697 and 6,539,265 bp.
223 Actually this gene is reported to be fixed for the p.Cys313Tyr variant in the Piedmontese breed
224 (Casas et al. 1999). Even though this causative mutation is not present in the SNP chip, the signal
225 has been detected in the adjacent markers.
226 A rather unexpected result has been obtained on BTA14, where no relevant signals in the
227 region where the *DGAT1* locus (*dyacylglycerol O-acyltransferase 1*) maps were found. However, it
228 should be remembered that some studies have reported the fixation of the p.Lys232Ala variant both
229 for the Italian Brown and Piedmontese breeds (Kaupe et al., 2004).
230 Other detected genomic regions, in agreement with previous researches on selection
231 signatures in cattle, were those harbouring genes affecting coat colour. These loci have been under
232 strong selection considering the importance of this trait in defining cattle breeds (Flori et al., 2009;
233 Wiener and Wilkinson 2011). In the present study, two selection signatures were observed on
234 BTA18 (12-13Mbp) and between 72-73 Mbp still on BTA6. In these chromosomic regions are
235 located the *Melanocortin 1 receptor (MC1R) and* the *Kit* (*V-kit Hardy-Zuckerman 4 feline sarcoma*
236 *viral oncogene homolog),* loci respectively. The Extension locus controls melanine synthesis. The
237 presence of three different alleles (E, E^1 and e) in both cattle breeds considered in this study has
238 been reported (Russo et al., 2007). The *Kit* locus is responsible for the "Piebald" spotted coat-colour
239 pattern in cattle and other species. This is interesting because Brown Swiss and Piedmontese breeds
240 did not show Piebald phenotype (Stella et al. 2010), confirming the complex genetic architecture of

241 coat colour in mammals.

242 In the present study, strong selection signals have been identified also in genomic regions
243 not previously associated to traits of economic importance.

244 Several genes related to calcium homeostasis and metabolism were found. *Osteocrin*
245 (*OSTN*) on BTA1, the *calcitonin receptor (CALCR)* and *calmodulin 2 (CAM2)* on BTA4 encode for
246 bone specific proteins that appears to act as soluble osteoblast/osteoclast regulators (Thomas et al.,
247 2003). The analysis of BTA10 outliers has revealed the presence of one interesting gene, the
248 *GREM1* that encode for the *gremlin 1*, a protein required for the osteoblastic activity and mineral
249 apposition (Canalis et al. 2012). Moreover, on BTA15 *STIM1* (*Stromal interaction molecule 1*, was
250 highlighted. It is expressed in mammary gland and it is essential for the cellular storage of calcium
251 and the activation of the calcium influx pathway (Li et al., 2012). Actually, milk production is a
252 complex biological process involving different tissues and governed by many genes (Finucane et
253 al., 2007; Lemay et al., 2009). Bone and mammary tissues are related via the same signalling
254 pathways (Cohen at al., 2004). Bone is a dynamic tissue continually modelled through the
255 coordinate actions of the bone forming osteoblast and resorbing osteoclast (Budayr et al., 1989).
256 Lactation is considered one of the most important events that determine bone remodelling due to the
257 relevant calcium mobilization (Qing et al., 2012). During the production of milk, plasmatic Ca^{+2}
258 entries into the mammary epithelial cells through a phenomenon called calcium influx pathway
259 (McAndrew et al., 2011).

260 A further set of highlighted genomic regions are those were genes related to epithelial cell
261 proliferation, skeletal muscle and bone morphogenesis map. As reported in the previous section, a
262 selection signature in a large region spanning between 7 and 10 Mbp was found on BTA2. Several
263 genes involved in the biology of muscular apparatus have been mapped in this region. Examples are
264 the *Bridging integrator 1* (*BIN1*), that plays an important role in muscle cell biology (Sedwick

2010), and the *Solute Carrier family 40 (iron regulated transporter), member1* (*SLC40A1*) locus, that codes for the ferroportin 1 (FPN1) a protein with an essential role in the regulation of iron levels on the body.

On BTA11 the *Bone morphogenetic protein 10 (BMP10)* a growth factor belonging to the TGF-ß superfamily known for its ability to induce bone and cartilage development (Groenveld and Burger, 2000) was found. Moreover, *SNAI3 (Snail homolog 3*) and *CDH15 (cadeherin 15, type1, M-cadherin (myotubule)* were highlighted on BTA18 (12,908,122-13,260,964 bp). They are involved in the skeletal morphogenesis and myoblast differentiation (Moran et al., 2002; Zhuge et al. 2005). An evident peak around 26-27 Mbp was observed on BTA20. In this genomic region is annotated the *Follistatin* (*FST*) locus. This protein acts blocking the binding of *Myostatin* to its receptor and causing an abnormal muscle development (McPherron and Lee, 2001). Table 2 reports other regions identified by peaks exceeding Control Chart limits and the annotated genes involved in skeletal muscle development and metabolism. Strong selection signatures observed in regions of genes related to muscle development, differentiation and metabolism could be interpreted as signs of selection within the Piedmontese. However, it should be remembered that the Brown Swiss was originally a dual purpose breed. Thus some of these genes might have also contributed to determine the Brown's phenotype. A deeper knowledge of the role of these genes in muscular cells could be of help for selecting markers useful for beef cattle breeding.

In this study few putative candidate genes were detected for lipid metabolism (Table 2). This is probably due to the fact that intramuscular fat deposition not only depends on the genetic background but also by other factors such as age, sex, nutrition and farm conditions.

An interesting result was the identification of numerous putative candidate genes involved in the reproductive function (Table 2). Among them, the specific ligand (*KITLG*) for the *Kit* receptor was identified on BTA5. Actually the interaction between *kit* and its ligand is crucial for

289 fertility (Mithraprabhn and Loveland, 2009). Such results suggest a further deepening of the genetic
290 basis of relationships between production and fertility traits (Bello et al., 2012).

291 Finally, this genome wide analysis highlighted the presence of selection signatures for a
292 group of similar genes. Six genes belonging to the Calpains gene family were detected in four
293 different chromosomes: *CAPN 7* on BTA1, *CAPN 13* and *14* on BTA11, *CAPN 5* on BTA15, and
294 *CAPN 2* and *8* on BTA16, respectively. Several studies indicate calpains as regulators of apoptosis
295 and suggest an involvement of the calpain system during the muscle postmortem apoptotic pathway
296 (Mohanty et al., 2010). The interaction among calpains and other proteases is considered a
297 foundamental step for after slaughtering meat tenderization (Koohmaraie, 1992). A multi gene
298 family is formed by duplication of a single original gene. In cattle, 3.1% of the genome is
299 composed of duplicated genes, most of which encoding proteins involved in innate immunity,
300 sensory receptors and reproduction (Elsik et al., 2009). Generally, the expansion or contraction of
301 gene families can be due to chance or is the result of natural selection. Gene gain or loss are so
302 considered to be an incentive for evolutionary change and as a common advantageous response to
303 selective regimes (Demuth et al., 2006).

304 A detailed list of putative genes for all 29 bovine chromosomes highlighted by the Control
305 Chart outliers values is summarized in Table 2. All gene content information presented was derived
306 from the UCSC Genome Broswer Gateway (http://genome.ucsc.edu/) using the fourth draft of
307 bovine genome sequence assembly (Btau 4.0) and from NCBI or Swiss ProtK Source consultation.

308 Results obtained in this study on the comparison between two cattle breeds with different
309 production aptitude, beef and dairy, agree with previous report on milk QTL (Cohen et al., 2004)
310 and transcriptome analysis (Bionaz and Loor, 2008; Lemay et al., 2009). Moreover it confirmed
311 what observed in previous comparisons between cattle breeds (Flori et al., 2009; Stella et al., 2010;
312 Qanbari et al. 2011). Differences have been found with the QTL analysis carried out by Prasad et al.

(2008) on BTA19 and 29, where selection signatures in different chromosomal regions were found. A possible explanation could be represented in the different genomic assemblies used.

Conclusions

The combined use of a LOWESS regression and a Control Chart approach here proposed was effective in studying the genetic differences between the Piedmontese and the Italian Brown cattle breeds. In particular, the local regression was able to yield a smooth F_{st} pattern, easy to interpret compared to raw data. The Control Chart allowed for a quite simple detection of significant F_{st} values that may indicate selection signatures. The method was validated by comparing results obtained on several chromosomes with previous reports in cattle (Hayes et al. 2008a and b; Flori et al., 2009; Stella et al 2010; Qanbari et al. 2011). Moreover, some regions harboring genes not yet associated to traits of economic importance for livestock have been detected. In particular, genes involved in the calcium metabolism and muscle biology have been highlighted. The methodology could be proposed as an easy approach for performing a whole genome scan in studies aimed at identifying selection signatures by using high throughput SNP maps.

Authors' contributions

E.P., S.S. and N.P.P. M. planned and performed analysis and drafted the manuscript. A.A., G.G., C.D., R.S. and G.M., performed the analysis. All authors reviewed the manuscript.

Acknowledgments

This work was funded by the Italian Ministry of Agriculture (grants SELMOL and INNOVAGEN). Authors wish to acknowledge Proff. Alessio Valentini and Paolo Ajmone Marsan for providing bull

337 genotypes.

338

339 **References**

340 Andersson L., and Georges M. (2004) Domestic-animal Genomics: Deciphering the genetics of
341 complex traits. *Nature Reviews Genetics* **5,** 202-212.

342 Band M.R. (2000) An ordered comparative map of the cattle and human genomes. *Genome*
343 *Research* **10,** 1359-1368.

344 Barendse W., Harrison B.E., Bunch R.J., Thomas M.B. and Turner L.B. (2009) Genome wide
345 signatures of positive selection: The comparison of independent samples and the identification of
346 regions associated to traits. *BMC Genomics* **10,** 178.

347 Bello N.M., Stevenson J.S., Tempelman R,J. (2012) Invited review: Milk production and
348 reproductive performances: Modern interdisciplinary insights into an enduring axiom. *Journal of*
349 *Dairy Science* **95,** 5461_5475.

350 Bionaz M. and Loor J. (2008) Gene networks driving bovine milk fat synthesis during the lactation
351 cycle. *BMC Genomics* **9,** 366.

352 Biswas S. and Akey J.M. (2006) Genomic insights into positive selection. T*rends in Genetics* **22 ,**
353 437-446.

354 Budayr A.M., Halloran B.P., King J.C., Diep D., Nissenson R:A., Strewler G.j. (1989) High levels
355 of a parathyroid hormone-like protein in milk. *Proceedings of the National Academy of Science*
356 *U.S.A.* **86,** 7183-7185.

357 Canalis E., Parker K., Zanotti S. (2012) Gremlin 1 is required for skeletal development and
358 postnatal skeletal homeostasis. *Journal of Cell Physiology* **227,** 269-277.

359 Cañón J., Alexandrino P., Bessa I., Carleos C., Carretero Y., Dunner S., Ferran N., Garcia D.,
360 Jordana J., Laloë D., Pereira A., Sanchez A., Moazami-Goudarzi K. (2001) Genetic diversity

361 measures of local European beef cattle breeds for conservation purposes**.** *Genetic Selection*
362 *Evolution* **33,** 311-332.
363 Casas E., Keele J.W., Fahrenkrug S.C., Smith T.P., Cundiff L.V. and Stone R.T. (1999) Quantitative
364 analysis of birth, weaning, and yearling weights and calving difficulty in Piedmontese crossbreds
365 segregating an Inactive Myostatin Allele. *Journal of Animal Science* **77,** 1686-1692.
366 Ciampolini R., Moazami-Goudarzi K., Vaiman D., Dillmann C., Mazzanti E., Foulley J.L., Leveziel
367 H. and Cianci D. (1995) Individual multilocus genotypes using microsatellite polymorphisms to
368 permit the analysis of the genetic variability within and between Italian beef cattle breeds. *Journal*
369 *of Animal Science* **73,** 3259-3268.
370 Cleveland W.S. (1979) Robust Locally Weighted Fitting and Smoothing Scatterplots. *Journal of the*
371 *American Statistical Association*, **74,** 829-836.
372 Cohen M., Reichenstein M., Everts-van der Wind A., Heon-Lee J., Shani M., Lewin H.A., Weller
373 J.I., Ron M. and Seroussi E. (2004) Cloning and characterization of FAM13A1 a gene near a milk
374 protein QTL on BTA6: evidence for population-wide linkage disequilibrium in Israeli Holstein.
375 *Genomics* **84,** 374-383.
376 Cohen R.A. (1999) An introduction to PROC LOESS for local regression. SAS Institute Inc. Cary,
377 North Carolina, USA.
378 Cohen-Zinder M., Seroussi E., Larkin D.M., Loor J.J., Everts-van der wind A., DrackleyJ.K., Band
379 M.R., Hernandez A.G., Shani M., Lewin H.A., Weller J.I., Ron M. (2005) Identification of a
380 missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6
381 affecting milk yield and composition in Holstein cattle. *Genome Research* **15,** 936-44.
382 Coskun A., Serster M., Fraterman A., Unsal I. (2008) a new international quality control chart based
383 on biological variation. *Accreditation and Quality Assurance: Journal for Quality, Comparability*
384 *and Reliability in Chemical Measurement* **13**, 69-75

385 Das S., Duggal P., Roy R., Myneedu V.P., Behera D., Prasad H.k., Bhattacharya A. (2012)
386 Identification of hot and cold spots in genome of *Mycobacterium tubercolosis* using Shewart
387 Control Chart. *Nature Scientific Reports* **2,** article number:297 doi:10.1038/srep00297.
388 Demuth J.P., DeBie T., Stajich J.E., Cristianini N., Hahm M.W. (2006) The Evolution of
389 mammalian gene families. *PLoS ONE* **1,**: e85. doi.10.B71/journal.pone.0000085.
390 Deng L.B., Tang X.L., Kang J., Wang Q.Y., Zeng C.Q. (2007) Scanning for signature of
391 geographically restricted selection based on population genomics analysis. *Chinese Science Bulletin*
392 **52,** 2649-2656.
393 Diskin S., Li M., Yang S., Glesseen J., Hakonarson H., Bucan M., Maris J.M., Wang K. (2008)
394 Adjustment of genomic waves in signal intensities from whole- genome SNP genotyping platforms.
395 *Nucleic Acid Research* **36**, e126.
396 Doebley J.F., Gaut B.S. and Smith B.D. (2006) The Molecular Genetics of Crop Domestication.
397 *Cell* **127,** 1309-1321.
398 Ellis A.D. (2010) The impact of nutrition on the health and welfare of horses. *EAAP* **128,** 53-74
399 Wageningen Academic Publishers. The Netherlands.
400 Elsik C.G., Tellam R.S., Worley K.C., and the Bovine Genome Sequencing and Analysis
401 Consortium (2009) The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and
402 Evolution. Science **324,** 522-528.
403 Ferencakovic M., Hamzic E., Gredler B., Solberg T.R., Klemetsdal G., Curik I., Solkner J. (2012)
404 Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected
405 cattle populations. *Journal of Animal Breeding and Genetics* **ISS0931-2668,** 1-8.
406 Finucane K.A., McFadden T.B., Bond J.P., Kennelly J.J., Zhao F-Q. (2008) Oneset of lactation in
407 the bovine mammary gland: gene expression profiling indicates a strong inhibition of gene
408 expression in cell proliferation. *Functional Integrative Genomics* doi 10.1007/s10142-008-0074-y.

409 Flori L., Fritz S., Jaffrézic F., Boussaha M., Gut I., Heath S., Foulley J.L., and Gautier M. (2009)
410 The Genome Response to Artificial Selection: A Case Study in Dairy Cattle. *PloS One* **4**, e6595.
411 Gibson J., Morton N.E., Colins A. (2006) Extended tracts of homozygosity in outbred human
412 populations. *Human Molecular Genetics* **15,** 789-795.
413
414 Groenveld E.H.J and Burger E.H. (2000) Bone morphogenetic proteins in human bone regulation.
415 *European Journal of Endocrinology* **142,** 1-9.
416 Hayes B.J., Chamberlain A.J., Maceachern S., Savin K., McPartlan H., MacLeod I. (2008a) A
417 genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef
418 cattle. *Animal Genetics* **40,** 176–184.
419 Hayes B.J., Lien S., Nilsen H., Olsen H.G., Berg P. (2008b) the origin of selection signatures on
420 bovine chromosome 6. *Animal Genetics* **39,** 105-111.
421 Jacoby W.G. (2000) Loess: a nonparametric, graphical tool for depicting relationships between
422 variables. *Electoral Studies* **19,** 577-613.
423 Kaupe B., Winter A., Fries R., Erhardt G. (2004) DGAT1 polymorphism in *Bos Indicus* and *Bos*
424 *Taurus* cattle breeds. *Journal of Dairy Research* **71,** 182-187.
425 Khatkar M.S., Tier B., Hobbs M., Kharkar D., Cavanagh J.A.L. Crump R., Moser G., Raadsma
426 H.W. (2010) Genome structure in Australian Holstein Frisian cattle revealed by combined analysis
427 of three high density SNP panels. *Proceeding Association Advancement Animal Breeding and*
428 *Genetics* **18,** 247-250.
429 Kijas J.W., Lenstra J.A., Hayes B., Boitard S., Porto Neto L.R., San Cristobal M., Servin B.,
430 McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan L.,
431 Dalrymple B., and International Sheep Genomics Consortium (2012) Genome-Wide Analysis of the
432 World's Sheep Breeds Reveals High Levels of Historic Mixture and Strong Recent Selection. *PloS*

433 *Biology* **10,** e 10001258.

434 Koomharaie M. (1992) The role of (Ca^{2+})-dependent proteases (calpains) in post-mortem
435 proteolysis and meat tenderness. *Biochimie* **74,** 239-245.

436 Lemay D.G., Lynn D., Martin W.F., et al. (2009) The bovine lactation genome: insigths into the
437 evolutionof mammalian milk. *Genome Biology* **10,** R43 1-18.

438 Li T., Finch E.A., Graham V., Zhang Z.S., Ding J.D., Burch J., Oh-hora M., Rosenberg P. (2012)
439 STIM1-(Ca^{2+}) signalling is required for the hypertrophic growth of skeletal muscle in mice.
440 *Molecular Cell Biology* **15,** 3009-3017.

441 Mäki-Tanila A., Fernandez J., Toro M., and Meuwissen T. (2010) Assesment and management of
442 genetic variation. Local Cattle Breeds in Europe. Development of policies and strategies for self-
443 sustaining breeds. Wageningen Academic Publishers 98-117.

444 Maynard-Smith J. and Haigh J. (1974) The hitch-hiking effect of a fovourable gene. *Genetical*
445 *Research Cambridge* **23,** 23-35.

446 McAndrew D., Grice D.M., Peters A.A., Davis F.M., Stewart T., Rice M., Smart C., Brown M.A.,
447 Kenny P.A., Roberts-Thomson S.J., Monteith G.R. (2011) ORAI1- Mediated Calcium influx in
448 lactation and breast cancer. *Molecular Cancer Therapeutics* **10,** 448-460.

449 McPherron A.C. and Lee S.J. (2001) Regulation of myostatin activity and muscle growth.
450 *Proceedings of the National Academy of Science U.S.A.* **98**, 9306-9311.

451 Mithraprabhu S. and Loveland K.L. (2009) Control of KIT signalling in male germ cells: what can
452 we learn from other systems? *Reproduction* **138,** 743-757.

453 Mohanty T.R., Park K.M., Pramod A.B., Kim J.H., Choe H.S., Hwang I.H. (2010) Molecular and
454 Biological factors affecting skeletal muscle cells after slaughtering and their impact on meat quality.
455 *Journal of Muscle Foods* **21,** 51-78.

456 Moran J.L., Li Y.,Hill A.A., Mounts W.M:, Miller C.P. (2002) Gene expression changes during

457 mouse skeletal myoblast differentiation revealed by transcription profiling. *Physiological Genomics*
458 **10,** 103-111.
459 Morrison L. W. (2008) The Use of Control Chart to Interpret Environmental Monitoring Data.
460 *Natural Areas Journal* **28,** 66-73.
461 Prasad A., Schnabel R.D., McKay S.D., Murdoch B., Stothard P., Kolbehdari D., Wang Z., Taylor
462 J.F., Moore S.S. (2008) linkage disequilibrium and signature of selection on chromosomes 19 and
463 29 in beef and dairy cattle. *Animal Genetics* **39,** 597-605.
464 Przeworski M., Graham C., and Wall J.D. (2005) The Signature of Positive Selection on Standing
465 Genetic Variation. *Evolution* **59**, 2312-2323.
466 Purfield D.C., Berry D.P., McParland S., Bradley D.G. (2012) runs of Homozygosity nd population
467 history in cattle. *BMC Genetics* **13,** 70.
468 Qanbari S., Gianola D., Hayes B., Schenkel F., Miller S., Moore S., Thaller G., Simianer H. (2011)
469 Application of site haplotype-frequency based approaches for detecting selection signatures in
470 cattle. *BMC Genomics* **12**, 318.
471 Qanbari S., Pimentel E.C.G., Tetens J., Thaller G., Lichtner P., Sharifi A.R., Simianer H. (2010) A
472 genome-wide scan for signatures of recent selection in Holstein cattle. *Animal Genetics* **41,** 377–
473 389.
474 Qing H., Ardeshirpour L., Pjevic P.D., Dusevich V., Jahn K., Kato S., Wyslomerski J., bonewald
475 L.F. (2012) Demonstration of osteocytic perilacunar/canalicular remodeling in mice during
476 lactation. *Journal of Bone and Mineral Research* **27,** 1018-1029.
477 Questa-Marcos A., Szucs P., Close T.J., Flichking T., Muehlbauer G., Smith K.P., Hayes P.M. (2010)
478 Genome-wide SNPs and re-sequencing of growth habit and inflorescence genes in barley:
479 implications for association mapping in germplasm arrays varying in size and structure. *BMC*
480 *Genomics* **11,** 707.

Ron M. and Weller J.I. (2007) From QTL to QTN identification in livestock- winning points rather than knock-out: a review**.** *Animal Genetics* **38,** 429-439.

Russo V., Fontanesi L., Scotti E., Tazzoli M., Dall'Olio S., Davoli R. (2007) Analysis of melanocortin 1 receptor (MC1R) gene polymorphisms in some cattle breeds.their usefulness and application for breed traceability and authentication of Parmigiano Reggiano cheese. *Italian Journal of Animal Science* **6,** 257-272.

Schwarzenbacher H., Dolezal M., Flisikowski K., Seefried F., Wurmser C., Schlötterer and Fries R. (2012) Combining evidence of selection with association analysis increases to detect regions influencing complex traits in dairy cattle. *BMC Genomics* **13**, 48.

Sedwick C. (2010) BIN1: A protein with Great Heart. *PLoS Biology* **8,** e1000311.

Sheehy P.A., Riley L.G., Raadsma H.W., Williamson P., and Wynn P.C. (2009) A functional genomics approach to evaluate candidate genes located in a QTL interval for milk production traits on BTA6. *Animal Genetics* **40,** 492-498.

Stella A., Ajmone-Marsan P., Lazzari B., Boettcher P. (2010) Identification of Selection Signatures in Cattle Breeds Selected for Dairy Production. *Genetics* **185,** 1451-1461.

Stephan W., Song Y.S., Langley C.H. (2006) The hitchhiking effect on linkage disequilibrium between linked neutral loci. *Genetics* **172,** 1-37.

The Bovine HapMap Consortium (2009) Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* **324,** 528-531.

Thomas G., Moffatt P., Salois P., d M.H., Gingras R., Godin E., Miao D., Goltzman L., Lanctot C. (2003) Osteocrin a novel bone-specific secreted protein that modulates the osteoblast phenotype. *Journal of Biological Chemistry* **278,** 50563-50571.

Weikard R., Widmann P., Buitkamp J., Emmerling R. and Kuehn C. (2012) Revisiting the quantitative trait loci for milk production traits on BTA6. *Animal Genetics* **43**, 318-323.

505 Weir B.S. and Cockerham C.C. (1984) Estimating F-Statistics for the Analysis of Population
506 Structure. *Evolution* **38**, 1358-1370.
507 Weir B.S., Cardon L.R., Anderson A.D., Nielsen D.M. and Hill W.G. (2005) Measures of human
508 population structure show heterogeneity among genomic regions. *Genome Research* **15,** 1468-1476.
509 Westgard J.O., Barry P.L., Hunt M.R. (1981) A Multi- Rule Shewart Chart for Quality Control in
510 Clinical Chemistry. *Clinical Chemistry* **27,** 493-501.
511 Wiener P. and Wilkinson S (2011) Deciphering the genetic basis of animal domestication.
512 *Proceedings of The Royal Society B* **doi:10.1098/rspb.2011.1376**
513 Zhuge **X**, Kataoka H, Tanaka M, Murayama T, Kawamoto T, Sano H, Togi K, Yamauchi R, Ueda Y,
514 Xu Y, Nishikawa S, Kita T, Yokode M. (2005) Expression of the novel Snai-related zinc-finger
515 transcription factor gene Smuc during mouse development. *International Journal of Molecular*
516 *Medicine* **15,** 945-948.
517
518
519
520

521 Table 2 List of putative candidate genes obtained on the basis of Control Chart outliers

Biological Function	BTA	Position Mbp	Gene name
Immune Response	5	81,763,516-81,779,866	*USP18 ubiquitin specific peptidase 18*
	16	23,622,572-23,625,154	*TLR5 toll-like receptor 5*
	17	57,084,217-57,115,368	*HVCN1 hydrogen voltage-gated channel 1*
	18	1,880,236-12,887,173	*CYBA cytochrome b-245, alpha polypeptide*
	19	21,395,686-21,409,196	*TMIGD1 transmembrane and immunoglobulin domain containing 1*
	26	23,471,864-23,478,382	*NFKB2 nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 (p49/p100)*
Reproduction	1	155,943,716-155,956,150	*EAF1 ELL-associated factor 1*
	3	86,007,282-86,200,728	*AK4 Adenylate kinase 4*
	5	20,587,724-20,612,963	*KITLG Kit ligand*
	6	37,961,724-37,987,164; 38,153,046-38,199,153; 38,227,954-38,378,385	*LAP3 leucine aminopeptidase 3;* *NCAPG non-SMC condensing I complex, subunit G;* *LCORL ligand dependent nuclear receptor corepressor-like*
	8	104,876,401-104,908,801	*TXNDC8 Thioredoxin domain containing 8 (spermatozoa)*
	9	41,225,543-41,246,855	*AMD1 adenosylmehtionine decarboxylase 1*
	10	36,873,000-36,890,219	*TYRO3 TYRO3 protein tyrosine kinase*
	11	70,119,086-70,174,863	*GMCL1 germ cell-less, spermatogenesis associated 1*
	14	60,023,782-60,033,403	*ODF1 outer dense fiber of sperm tails 1*
	15	18,520,179-18,520,292; 24,046,636-24,397,152	*FDX1 ferredoxin 1;* *CADM1 cell adhesion molecule 1*
	18	13,648,996-13,652,641	*SPATA2L spermatogenesis associated 2-like*
	19	24,498,808-24,501,792; 24,628,862-24,646,107; 50,216,969-50,223,538	*GSG2 germ cell associated 2 (haspin);* *P2RX1 purinergic receptor P2X, ligand-gated ion channel,1 ;* *DDX5 DEAD (Asp-Glu-Ala-Asp) box helicase 5*
Cell growth, proliferation and differentiation	1	76,216,039-76,832,685	*FGF12 fibroblast growth factor 12*
	12	70,974,850-71,682,818	*HS6ST3 heparan-sulfate 6-O-sulfotransferase 3*
	13	47,627,052-47,683,993 48,488,115-48,536,904	*CDS2 CDP-diacylglycerol synthase (phosphatidate cytydyltransferase) 2 ;* *FERMT1 fermitin family member1*
	14	60,169,396-60,307,900	*UBR5 ubiquitin protein ligase E3*

			component n-recognin 5
	16	20,839,081-20,931,656 26,878,683-26,905,046	*TGFB2 transforming growth factor, beta2;* *PSEN2 presenilin 2 (Alzheimer disease 4)*
	17	57,146,787-57,165,849	*PPP1CC protein phosphatase 1, catalytic subunit, gamma isozyme*
	19	*35,124,710-35,129,750* *35,535,495-35,544,295* *35,953,771-35,969,817* *45,567,703-45,574,688*	*MAPK7 mitogen- activated protein kinase 7;* *DRG2 developmentally regulated GTP binding protein 2;* *FLCN folliculin;* *GRN granulin*
Ions metabolism	1	78,466,667-78,488,928	*CLDN16 claudin 16*
	2	9,451,265-9,580,452	*CALCRL calcitonin receptor- like*
	4	11,016,143-11,126,171	*CALCR calcitonin receptor*
	5	81,136,111-81,146,812	*KCTD17 potassium channel tetramerisation domain containing 17*
	8	*11,700,825-11,763.811*	*ACO1 aconitase 1, soluble*
	17	*56,466,582-56,498,348,;* *56,790,348-56,488,450*	*CAMKK2 calcium/calmodulin-dependent protein kinase kinase 2, beta;* *ATP2A2 ATPase, Ca++ transporting, cardiac muscle, slow twitch 2*
	18	47,597,196-47,605,452	*KCNK6 potassium channel, subfamily k,member 6*
	19	56,790,348-56,844,450; 24,594,778-24,623,204	*ATP2A3 ATPase, Ca++ transporting, ubiquitous;* *CAMKK1 calcium/calmodulin-dependent protein kinase kinase 1, alpha*
	24	31,254,115-31,532,051	*KCTD1 potassium channel tetramerisation domain containing 1*
	26	22,854,587-22,857,882; 24,558,695-24,564,440	*KCNIP2 Kv channel interacting protein 2;* *CALHM3 calcium homeostasis modulator 3*
Lipid metabolism	2	6,192,072-6,348,621	*HIBCH 3 hydroxibutirril o idrolase*
	5	53,700,174-53,700,270	*ACAT2 AcetylCoA acetyltransferase 2*
	10	59,440,432-59,504,627	*CYP19A1 cytochrome P450, family 19, subfamily A, polypeptide1*
	13	48,423,438-48,446,513	*CRLS1 cardiolipin synthase 1*
	15	55,827,654-56,160,380	*ACER3 alkaline ceramidase 3*
	18	13,212,190-13,250,827	*ACSF3 Acyl-CoA synthase family member 3*
	19	35,671,152-35,687,188	*SREFB1 sterol regulatory element*

			bibdingtranscription factor 1
mammary gland metabolism	2	10,226,975-10,322,817	*ITGA V integrin alpha V*
	6	37,351,167-37,421,683 37,431,966-37,490,645 37,511,673-37,518,636 72,298,906-72,346,677 72,741,252-72,828,528	*ABCG2 ATP-binding cassette, sub-family G, (WHIT), member 2;* *PKD2 polycistic kidney disease 2;* *SPP1 secreted phosphoprotein 1;* *PDGFRA platlet-derived growth factor receptor, alpha poypetide;* *KIT V-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog*
	7	62,635,246-62,657,995	*SPARC secreted protein,acidic, cysteine-rich (osteonectin)*
	10	29,529,387-29,541,874 36,595,794-36,596,071	*GREM1 gremlin1, DAN family BMP antagonist;* *IGFBP3 insulin-like growth factor binding protein 3*
	14	12,506,878-12,583,201	*MTMR2 myotubularin related protin 2;*
	15	20,478,802-20,482,029 50,442,087-50,753,021	*CRYAB crystalline alpha B;* *STIM1 stromal interaction molecule 1*
	18	14,699,407-14,998,970	*ITGF1 integrin alpha FG-GAP repeat containing 1*
	19	35,122,081-35,124,619 35,823,315-35,854,048	*MFAP4 microfibrillar associated protein 4;* *PEMT phosphatidylethanolamine N-methyltransferase*
	24	30,845,569-30,860,104	*AQP4 aquaporin 4*
	27	48,475,540-48,478,931	*OXSM 3-oxyacyl-ACP synthase, mitochondrial*
bone and muscle metabolism	1	77,682,355-77,718,578 155,717,664-155,777,449	*OSTN osteocrin;* *CAPN7 calpain 7*
	2	5,595,799-5,652,801 6,532,697-6,539,265 7,066,569-7,148,685 7,740,061-7,779,695	*BIN1 bridging integrator1;* *MSTN myostatin;* *SLC40A1 solute carrier family 40(iron regulated transporter) member1;* *COL3A1 collagen type (III) alpha 1*
	7	5,824,715-5,935,402	*MYO9B myosin IXB*
	8	11,291,512-11,308,875 105,221,050-105,315,564	*CLU clusterin* *MUSK muscle, skeletal, receptor tyrosine kinase*
	10	19,387,377-19,414,041	*PKM pyruvate kinase, muscle*
	11	69,145,567-69,152,285 70,648,036-70,648,340 71,029,777-71,105,164	*BMP10 bone morphogenetic protein 10;* *CAPN14 calpain 14;* *CAPN13 calpain 13*

	13	48,488,115-48,536,904	*FERMT1 fermitin family member 1;*
	15	11,852,140-11,854,278 56,045,818-56,103,271	*PPP1R14C protein phosphatase 1, regulatory (inhibitor) subunit 14C ;* *CAPN5 calpain 5*
	16	24,021,217-24,065,788 24,108,032-24,166,355	*CAPN8 calpain 8;* *CAPN2 calpain 2*
	17	56,905,068-56,915,878 57,330,762-57,338,500	*ARPC3 actin related protein 2/3 complex, subunit3 21kDa;* *MYL2 myosin light chain 2,regulatory, cardiac, slow*
	18	12,908,122-12,913,750 13,260,964-13,279,948 47,527,738-47,531,970 47,701,775-47,875,177	*SNAI3 snail homolog 3;* *CDH15 cadherin 15, type1,M-cadherin (myotubule);* *PPP1R14A protein phosphatase 1, regulatory (inhibitor) subunit 14A;* *RYR1ryanodine receptor 1(skeletal)*
	20	23,624,160-23,688,918 27,297,146-27,302,564	*GPBP1 GC-rich promoter binding protein 1;* *FST follistatin*
	21	45,895,690-45,898,343	*CFL2 cofilin 2 (muscle)*
	26	12,908,235-12,917,607 23,540,685-23,557,026	*ANKRD1 ankyrin repeat domain 1 (cardiac muscle);* *ACTR1A ARP1 actinn related protein 1 homolog a, centractyn alpha (yeast)*
others	10	19,817,179-19,849,769	*ADPGK ADP-dependent glukonidase*
	11	68,612,764-68,639,385 69,642,777-69,707,857	*CNRIP1 cannabinoid receptor interacting protein 1;* *GFTP1 glutamine-fructose-6 phosphate transaminase 1*
	15	20,576,533-20,611,864	*DLAT dihydrolipoamide S-acetyltransferase*
	18	13,776,888-13,778,639	*MC1R melanocortin 1 receptor (alpha melanocyte stimulating hormone receptor)*
	19	45,226,420-45,227,150 45,325,106-45,329,822	*PPY pancreatic polypeptide* *G6PC3 glucose 6 phosphatase, catalytic, 3*

Captions to figures:

Figure 1 Comparison of average heterozygosity (Hobs) per chromosome between the two breeds (black = Piedmontese, grey = Italian Brown).

Figures 2. Pattern of raw F_{st} data calculated for SNP located along the BTA 6 (a) and 23 (b); predicted F_{st} values for the SNP located along BTA6 (c) and 23 (d) using the LOWESS regression with a smoothing parameter of 0.009 and 0,021 respectively; Control Chart of predicted F_{st} values for BTA6 (e) and 23 (f). Solid line: Mean, dotted lines are: upper control limit (UCLI) and lower control limit (LCLI).

Figure 3 Plot of comparison between Sliding Windows versus LOWESS on BTA6. Solid line: Sliding Windows method, dotted line: LOWESS methodology